

# Trends in molluscan gene sequence similarity: An observation from genes expressed within the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791) (Neogastropoda: Muricidae)

Patrick W. Laffy  
Kirsten Benkendorff<sup>1</sup>  
Catherine A. Abbott

School of Biological Sciences  
Flinders University  
GPO Box 2100 Adelaide  
SA 5001 AUSTRALIA

## ABSTRACT

This study investigates the phylogenetic distribution of homology to *Dicathais orbita* hypobranchial gland genes based on tBLASTx pairwise sequence alignments from the Genbank database. Suppressive subtractive hybridization was used to obtain 417 non-redundant genes that were up-regulated or uniquely expressed in the hypobranchial gland relative to mantle tissue. Of these, 133 sequences revealed matches to the database with the remaining 68% of genes appearing as apparently novel sequences. Homologous sequence matches were observed for a wide range of evolutionarily divergent taxa, encompassing animals, protozoans, plants, fungi, bacteria, and viruses. The highest frequency of homology was found towards chordate sequences, followed by the Mollusca, which highlights the current bias in availability of vertebrate versus invertebrate sequences in the database. An unexpectedly high proportion of matches were also found toward the Ciliophora, indicating a possible symbiotic relationship, as well as the Ascomycota and Streptophyta, which share the ability to biosynthesize indole derivatives with Muricidae such as *Dicathais orbita*. Overall, these results reveal the usefulness of undertaking sequence comparisons in gene expression and highlight the current paucity of knowledge of molluscan genomes.

*Additional keywords:* Gastropoda, DNA

## INTRODUCTION

The development of genomic technologies has had a dramatic effect on all fields of biological sciences (Collins et al., 2003). Since the completion of the human genome project in 2003 (Collins et al. 2003), the number of genomes available has grown dramatically. As of November 2007, a total of 426 eukaryotic (24 complete,

164 undergoing assembly and 238 in progress) and 599 bacterial genomes were available on the Genbank database (NCBI, 2007). The increased number of genomes available enhances our understanding of the biology of the species in question and provides a basis for comparative studies in functional biology. Despite this increase in data, trends in comparative genomics favor the analysis of mammalian sequences (Barnes et al. 2004), and often the homologous identification and classification of non-vertebrate sequences is more challenging.

The Mollusca has been identified as the second most diverse and speciose phylum in the animal kingdom, with members present in marine, freshwater, and terrestrial environments (Pechenik, 2000). Despite their abundance and the economic importance of many species (Beesley et al., 1998), the genome of mollusks remains relatively uncharted. So far, the complete genome has only been sequenced for the Californian sea hare *Aplysia californica* Cooper, 1863, and this is yet to be annotated (NCBI, 2007). The bivalves *Argopecten irradians* (Lamarck, 1819), *Crassostrea virginica* (Gmelin, 1791), and *Spisula solidissima* (Dillwyn, 1817), which are all important fisheries resources, and the medically important freshwater snail *Biomphalaria glabrata*, are currently undergoing sequencing (NCBI, 2007). Nevertheless, a major hurdle in molluscan genomics lies in defining the functions of sequences identified. Sequence homology has been used heavily to assign functions in mammalian genomes. However, the lack of currently available molluscan and invertebrate sequences limits the ability to assign gene function using comparative classifications drawn from existing invertebrate genomic sequence information. Nevertheless, broader comparisons to more distantly related organisms could yield novel information about well conserved genes or genes that have independently evolved convergent functions in distinct taxa.

The hypobranchial gland of neogastropods is a uniquely molluscan organ (Beesley et al., 1998) of uncertain

<sup>1</sup> Author for correspondence: kirsten.benkendorff@flinders.edu.au

origin and function (Westley et al., 2006). Within the family Muricidae, it is the well known source of the ancient dye Tyrian purple (Baker, 1974; Cooksey, 2001). Tyrian purple is generated by a series of chemical reactions from indoxyl sulphate precursors that are brominated secondary metabolites thought to be derived from the amino acid tryptophan (Westley et al., 2006). While the Muricidae are thought to be the only source of the purple brominated dye, the related blue dye indigo is produced by a number of other taxa including plants, bacteria and fungi (Epstein et al., 1969; Meijer et al., 2006; Mayser et al., 2007). This presents an interesting case of apparent convergent evolution in biosynthetic capabilities.

Basic Local Alignment Search Tool (BLAST) analysis is a key tool used to identify orthologous genes from different organisms, and its use has been instrumental in classifying countless sequences (Galagan et al., 2003; Venter et al., 2001). The taxonomic classifications of high scoring BLAST matches with unclassified sequences are useful in identifying sequences with specific or variable functions and may indicate key gaps in the current sequence data for members of specific phyla. This study results from a larger project that is currently underway to identify the genes expressed in the hypobranchial gland of *Dicathais orbita* (Gmelin, 1791), a predatory marine gastropod belonging to the family Muricidae, order Neogastropoda. Here we report on our tBLASTx analysis, where sequences were translated into all possible protein translations and compared to all possible translations of every nucleotides sequence in Genbank, to observe trends in molluscan sequence similarity and assess the proportion of homologous genes expressed in this unique biosynthetic organ.

## MATERIALS AND METHODS

A suppressive subtractive hybridization (SSH) (Diatchenko et al., 1999) cDNA library containing the up-regulated and differentially expressed genes within the hypobranchial gland of *D. orbita*, when compared to mantle tissue gene expression, was created using a Clontech PCR-Select™ cDNA Subtraction Kit (Clontech, California, USA). The RNeasy® RNA extraction kit (Ambion, Texas, USA), TRI Reagent® (Ambion) and DNaseI (Invitrogen, CA, USA) digestion were used to obtain RNA from the hypobranchial glands and mantle of two *D. orbita* specimens. The subtraction was performed utilizing pooled hypobranchial gland transcripts as the tester population and pooled mantle transcripts as the driver population. Subtracted cDNA produced from SSH were cloned into pGEM®-T Easy vector (Promega, Wisconsin, USA). Colonies with inserts were selected, plasmid DNA was purified and sequencing was performed by Southpath and Flinders Sequencing Facility (Adelaide, Australia) or Australian Genome Research Facility (AGRF sequencing, Brisbane, Australia). A total of 554 plasmids were sequenced, and vector

sequence and adaptor regions were removed. Contigs were formed using Sequencher Version 4.1.4 yielding a non-redundant set of expressed sequence tags (ESTs) differentially expressed in the hypobranchial gland of *D. orbita*. In total, 417 unique resulting sequences were submitted to tBLASTx analysis and the highest scoring matches for all sequences with an e value smaller than  $1e^{-5}$  were collated. The phylum of the orthologous sequence was recorded, and in cases where the highest scoring tBLASTx matched a molluscan sequence, the class was determined. In cases where the matching sequence belonged to a member of the class Gastropoda, the family was also recorded. The total number of Genbank sequences for phyla with 5 or more sequence matches was recorded (Figure 1).

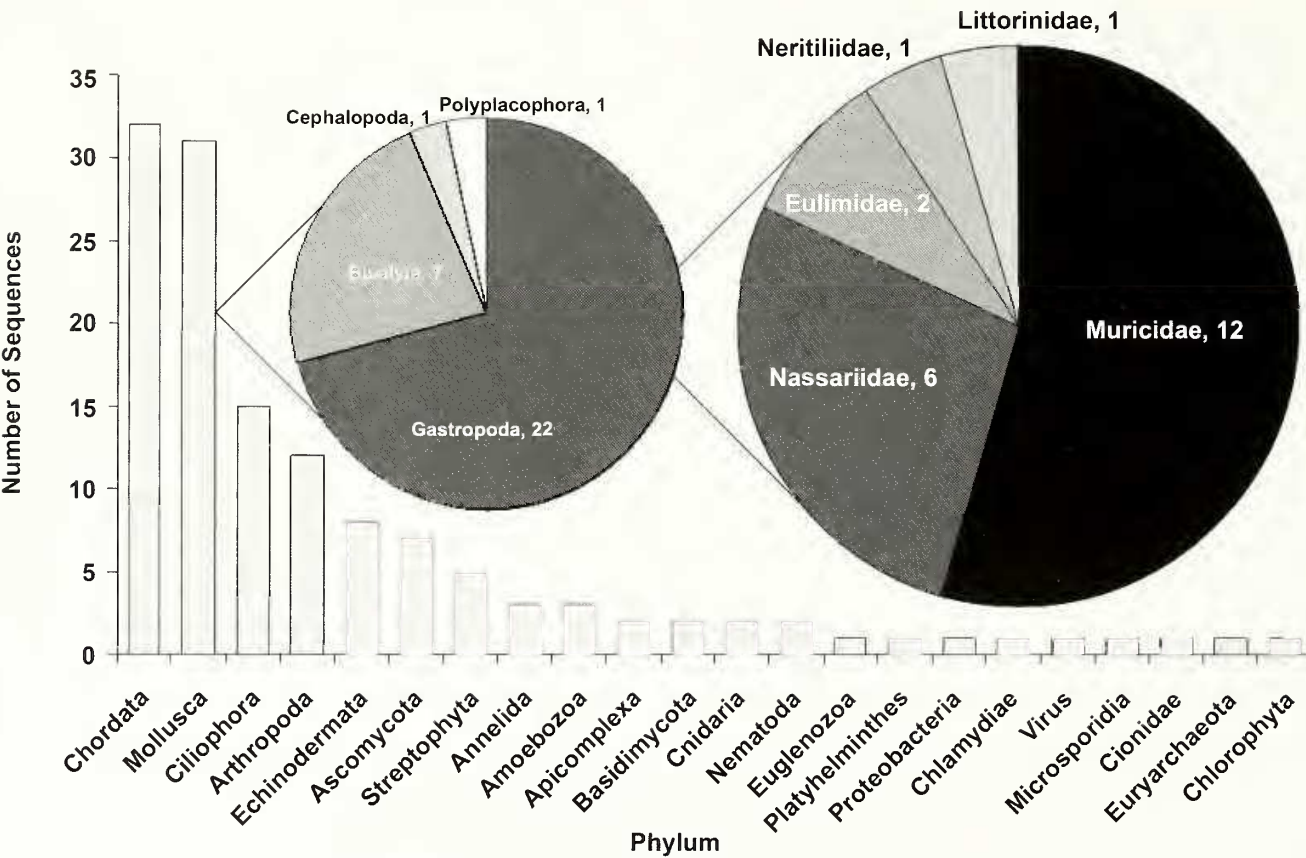
## RESULTS

A total of 133 sequences out of 417 (31.9%) resulted in significant tBLASTx matches, with 23 different phyla represented from the best scoring blast match for each identified sequence. Seven of these phyla had matches to 5 or more hypobranchial gland sequences from *D. orbita*. The Chordata showed the highest number of matches, with 32 homologous sequences identified, closely followed by the Mollusca with 31 matches (Figure 1). Ciliophora were the third most abundant phylum with 15 matches, followed by the invertebrate phyla Arthropoda and Echinodermata, with 12 and 8 sequences identified, respectively (Figure 1). There were seven Ascomycota homologs identified in *D. orbita*'s hypobranchial gland, as well as five from the Streptophyta (Figure 1).

Of the 31 molluscan sequence matches identified, 22 sequences matched gastropod sequences. Twelve of these gastropod sequence homologs belonged to other members of the Muricidae family (Figure 1). Further distribution of the sequence homology is detailed in Figure 1.

## DISCUSSION

While 133 of the sequences produced had BLAST matches that indicated the function of the transcripts, the remaining 284 genes sequenced from the hypobranchial gland of *Dicathais orbita* appear to be novel, highlighting the limited information currently available on molluscan genomes. The high frequency of matches to chordate sequences is likely to be due to the large abundance of vertebrate sequences in the public database (Barnes et al., 2004) (Table 1). There are currently over 57 million gene sequences from the Chordata, compared to less than 600,000 molluscan sequences available (Table 1). There is clearly a bias towards a high proportion of tBLASTx matches returning matches to human and other chordate sequences, which have over 90 times the number of molluscan genes available for sequence alignment.



**Figure 1.** Phyla represented by highest scoring tBLASTx matches of genes expressed in the hypobranchial gland of *Dicathais orbita*. A total of 417 non-redundant EST sequences were analysed using tBLASTx and the resulting 133 significant matches (E value < 10<sup>-5</sup>) were placed into 23 categories, based on the phylum grouping of the highest-scoring tBLASTx matches. Sequences grouped in the phylum Mollusca were further classified into the corresponding Class of the best tBLASTx match. Gastropod sequences were further divided according to family of the highest scoring tBLASTx matches.

The abundance of matches to sequences from the Ciliophora was unexpected, particularly since the number of ciliate sequences in public databases is just over 300,000 (Table 1). It is possible these protozoan gene matches actually result from ciliate genomes derived from endosymbionts occurring within the hypobranchial gland of *D. orbita*. Ciliates are ubiquitous protists that commonly form relationships with other species, such as the parasitic *Ichthyophthirius multifiliis* (Abernathy et al., 2007) and the symbiotic *Euplotes uncinatus* (Lobban et al., 2005).

The abundance of matches to arthropod species was not unexpected due to the shared ancestral relationship between the Mollusca and Arthropoda. However, the numerous matches to Echinodermata are less expected given that this phylum occurs on the deuterostome lineage along with chordates, which diverged from the mollusks and other protostomes over 100 million years ago (Heckman et al., 2001). Notably, there were relatively few matches to the Annelida (Figure 1) despite the fact that this abundant protostome phylum occurs within the Lophotrochozoan lineage alongside the Mollusca, which form a separate clade from the Ecdyzoa, including arthropods and nema-

todes (Aguinaldo et al., 1997). It is likely that the small number of annelid sequences available, less than 35,000 (Table 1), contributed to the small incidence of annelid sequence homology with our molluscan sequences. This further highlights the relatively limited genetic information that is available for so called "primitive" invertebrate phyla.

**Table 1.** Number of nucleotide sequences available on Genbank database for different phyla as published on the 18 December 2007. All data was compiled as published under the Taxonomy browser available on NCBI Entrez taxonomy home page <http://www.ncbi.nlm.nih.gov/sites/entrez?db=Taxonomy>.

Phylum	Genbank nucleotide sequences
Chordata	57,495,211
Mollusca	599,894
Ciliophora	303,304
Arthropoda	5,090,469
Echinodermata	941,561
Ascomycota	1,367,029
Streptophyta	22,413,399
Annelida	34,245



The frequency of sequence matches to the fungal Ascomycota and the plant Streptophyta was an unexpected finding. This is possibly related to the fact that members of both the Streptophyta and Ascomycota are capable of similar secondary metabolite production as is the muricid *Dicathais orbita*. Indigo is produced in *Isatis tinctoria* (phylum Streptophyta) (Epstein et al., 1967), and the production of indole compounds has been reported for *Candida glabrata* (phylum Ascomycota) (Mayser et al., 2007). These compounds are in the same chemical class of indole alkaloids as Tyrian purple, the brominated derivative of indigo secreted only from the hypobranchial gland of the Muricidae (Cooksey, 2001; Westley et al., 2006). These similarities in secondary metabolite production may influence the frequency of homology with genes expressed in the hypobranchial gland of *D. orbita*. Further analysis of the conserved genes could help reveal some key biosynthetic enzymes and/or processes. As SSH allows for amplification of only up-regulated or uniquely expressed genes in this instance, we would expect sequences involved in chemical and protein biosynthesis to be amplified. This demonstrates that it is important to consider the source of expressed genes when interpreting sequence homology.

Another key observation is the frequency and variation of molluscan gene matches observed from our tBLASTx analysis. As mentioned, a total of 31 molluscan sequence matches were identified, with 22 gastropod sequences, 12 of which belonged to the family Muricidae (Figure 1). This trend is expected as species within the same family are expected to show greater homology with our *D. orbita* sequences. The key limiting factor to the number of muricid and gastropod sequence matches is the limited amount of sequencing that has been performed on these groups, only 1994 Muricidae sequences have been published on the NCBI database as of November 2007 (NCBI 2007). The majority of sequences available for muricids are highly conserved genes involved in phylogenetic analysis such as ribosomal RNA (Colgan et al., 2007; Harasewych et al., 1997; Oliverio and Mariottini, 2001), cytochrome oxidase 1 (Colgan et al., 2007; Harasewych, et al., 1997) and histone H3 sequences (Colgan et al., 2007). The frequency of positive matches to *D. orbita* hypobranchial gland genes is likely to increase as a broader range of sequences from additional Muricidae and other gastropod species are made available on Genbank.

From tBLASTx analysis, we have identified the phylogenetic distribution of species that share homology with *Dicathais orbita* gene sequences. While less than 32% of sequences could be positively matched on the gene databases, 31 matches were found encompassing species from both invertebrates and vertebrates within the Animal Kingdom, as well as eukaryotic plants, protozoans, fungi, some prokaryotes and even viruses. Most matches pertain to chordate sequences, and this may be attributed to the abundance of these sequences within databases. Nevertheless, many of the sequences match other molluscan species and other invertebrate phyla,

likely due to the close evolutionary relationships leading to conserved genes. A significant proportion of sequences belong to ciliate protozoans, and it is unclear whether this is due to similarities between these protists and *D. orbita* or the addition of ciliate genes within our hypobranchial gland expressed genes. The limited number of molluscan gene matches from our dataset supports the need for a larger number of molluscan sequences to be identified and released, encompassing a broader range of functional genes. Only then will we be able to accurately view trends in gene expression within the hypobranchial gland of *D. orbita*.

## ACKNOWLEDGMENTS

We appreciated funding support from an anonymous philanthropic foundation. Patrick Laffy is supported firstly by a Flinders University Faculty of Science and Engineering Research scholarship, followed by a Flinders University Postgraduate Research scholarship. We would also like to thank the South Australian Partnership for Advanced Computing, for the use of their BLAST portal, Chantel Westley for her assistance in specimen dissection and Dr. Peter Speck for his assistance in manuscript editing.

## LITERATURE CITED

- Abernathy, J.W., P. Xu, P. Li, D.H. Xu, H. Kucuktas, P. Klesius, C. Arias, and Z.J. Liu. 2007. Generation and analysis of expressed sequence tags from the ciliate protozoan parasite *Ichthyophthirius multifiliis*. BMC Genomics 8: 1–10.
- Aguinaldo, A.M.A., J.M. Turbeville, L.S. Linford, M.C. Rivera, J.R. Garey, R.A. Raff, and J.A. Lake. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387: 489–493.
- Baker, J.T. 1974. Tyrian purple: an ancient dye, a modern problem. Endeavour (Oxford). 33: 11–17.
- Barnes, D.W., C.J. Mattingly, A. Parton, L.M. Dowell, C.J. Bayne, and J.N. Forrest. 2004. Marine organism cell biology and regulatory sequence discovery in comparative functional genomics 46: 123–137.
- Beesley, P.L., G.J.B. Ross, and A. Wells. 1998. Mollusca: The Southern Synthesis. Fauna of Australia, CSIRO Publishing, Melbourne, Part A, xvi + 563 pp., Part B, viii + pp. 565–1234.
- Colgan, D.J., W.F. Ponder, E. Beacham, and J. Macaranas. 2007. Molecular phylogenetics of Caenogastropoda (Gastropoda: Mollusca). Molecular Phylogenetics and Evolution 42: 717–737.
- Collins, F.S., M. Morgan, and A. Patrinos. 2003. The Human Genome Project: lessons from large scale biology. Science 300: 286–290.
- Cooksey, C.J. 2001. Tyrian purple: 6,6'-dibromoindigo and related compounds. Molecules 6: 736–769.
- Diatchenko, L., S. Lukyanov, Y.F.C. Lau, and P.D. Siebert. 1999. Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes, cDNA preparation and characterization. Methods In Enzymology 303: 349–380.

- Epstein, E., M.W. Nabors, and B.B. Stowe. 1967. Origin of Indigo of Woad. *Nature* 216: 547–549.
- Galagan, J.E., S.E. Calvo, K.A. Borkovich, E.U. Selker, N.D. Read, D. Jaffe, W. FitzHugh, L.-J. Ma, S. Smirnov, S. Purcell, B. Rehman, T. Elkins, R. Engels, S. Wang, C.B. Nielsen, J. Butler, M. Endrizzi, D. Qui, P. Ianakiev, D. Bell-Pedersen, M.A. Nelson, M. Werner-Washburne, C.P. Selitrennikoff, J.A. Kinsey, E.L. Braun, A. Zelter, U. Schulte, G.O. Kothe, G. Jedd, W. Mewes, C. Staben, E. Marcotte, D. Greenberg, A. Roy, K. Foley, J. Naylor, N. Stange-Thomann, R. Barrett, S. Gnerre, M. Kamal, M. Kamvysselis, E. Mauceli, C. Bickel, S. Rudd, D. Frishman, S. Krystofova, C. Rasmussen, R.L. Metzenberg, D.D. Perkins, S. Kroken, C. Cogoni, G. Macino, D. Catcheside, W. Li, R.J. Pratt, S. A. Osmani, C.P.C. DeSouza, L. Glass, M.J. Orbach, J.A. Berglund, R. Voelker, O. Yarden, M. Plamann, S. Seiler, J. Dunlap, A. Radford, R. Aramayo, D.O. Natvig, L.A. Alex, G. Manhaupt, D.J. Ebbole, M. Freitag, I. Paulsen, M.S. Sachs, E.S. Lander, C. Nusbaum, and B. Birren. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868.
- Harasewyeh, M.G., S.L. Adamkewicz, J.A. Blake, D. Saudek, T. Spriggs, and C.J. Bult. 1997. Neogastropod phylogeny: A molecular perspective. *Journal of Molluscan Studies* 63: 327–351.
- Heckman, D.S., D.M. Geiser, B.R. Eidell, R.L. Stauffer, N.L. Kardos, and S.B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* 293: 1129–1133.
- Lobban, C.S., L. Modco, F. Verni, and G. Rosati. 2005. *Euplotes uncinatus* (Ciliophora, Hypotrichia), a new species with zooxanthellae. *Marine Biology* 147: 1055–1061.
- Mayser, P., M. Wenzel, H.J. Kramer, B.L.J. Kindler, P. Spitelner, and G. Haase. 2007. Production of indole pigments by *Candida glabrata*. *Medical Mycology* 45: 519–524.
- NCBI [National Center for Biotechnology Information]. 2007. ENTREZ Genome Project Home Website, <http://www.ncbi.nlm.nih.gov/sites/entrez?db=genomeprj>, accessed 17 December 2007.
- Oliverio, M. and P. Mariottini. 2001. A molecular framework for the phylogeny of *Coralliophila* and related muricoids. *Journal of Molluscan Studies* 67: 215–224.
- Pechenik, J.A. 2000. *Biology of the Invertebrates*. Fourth Edition. McGraw-Hill Higher Education, Boston, 578 pp.
- Venter, J.C., M.D. Adams, E.W. Myers, P.W. Li, R. J. Mural, G.G. Sutton, H.O. Smith, M. Yandell, C.A. Evans, R.A. Holt, J.D. Gocayne, P. Amanatides, R.M. Ballew, D.H. Huse, J.R. Wortman, Q. Zhang, C.D. Kodira, X.H. Zheng, L. Chen, M. Skupski, G. Subramanian, P.D. Thomas, J. Zhang, G.L. Gabor Miklos, C. Nelson, S. Broder, A.G. Clark, J. Nadeau, V.A. McKusick, N. Zinder, A. J. Levine, R.J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasolo, M. Flanagan, L. Florca, A. Halpern, S. Hammenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V.D. Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A.E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T.J. Heiman, M.E. Higgins, R.-R. Ji, Z. Ke, K.A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G.V. Merkulov, N. Milshina, H.M. Moore, A.K. Naik, V.A. Narayan, B. Neclam, D. Nusskern, D.B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z.Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C., Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S.C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M.L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferreria, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C., Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkuch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Rombold, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Dcen, K. Wolfe, J. Zaveri, K. Zaveri, J.F. Abril, R. Guigo, M.J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yoosseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A.D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. The Sequence of the Human Genome. *Science* 291(5507): 1304–1351.
- Westley, C., K. Benkendorff, and K.L. Vine. 2006. A proposed functional role for indole derivatives in reproduction and the defense system of Muricidae (Neogastropoda: Mollusca). *Life in Progress* editions, Roscoff, France, pp. 31–44.